4К

УДК 004.932.751

Ю.В. Крак, Д.В. Шкільнюк
Інститут кібернетики ім. В.М. Глушкова НАН України, м. Київ, Україна
krak@unicyb.kiev.ua, dimonshk@gmail.com

Аналіз елементів дактильної жестової мови

Метою даного дослідження є розробка та реалізація алгоритмів розпізнавання дактильної жестової мови, що використовується при спілкуванні з людьми з вадами слуху. В роботі розширюється та деталізується схема класифікації окремих елементів дактильної мови, розглядаються та описуються методи побудови контуру елемента дактильної мови.

Вступ і постановка задачі

В наш час постає проблема щодо забезпечення комунікації людей з вадами слуху з іншими представниками суспільства, що потребує розробки систем моделювання та розпізнавання жестової мови. Зокрема, необхідно приділити увагу розпізнаванню елементів дактильної жестової мови, яка, в свою чергу, зображує букви алфавіту. Адже люди з вадами слуху використовують дактильну мову для відображення власних назв, імен, іншомовних наукових термінів, речовин і т.д.

Розробці систем для розпізнавання елементів жестової мови присвячено значну кількість робіт, зокрема у працях [1], [2] описано системи розпізнавання жестової мови при використанні спеціального апаратного забезпечення. Ряд дослідників запропонували систему, яка ідентифікує жести конкретної особи [3]. Крім того, більшість підходів спрямовані на розпізнавання окремих ознак жестів або розпізнавання жестів, які характеризуються лише напрямком руху руки [4], [5].

Разом з тим у процесі відтворення елементів жестової мови виникає ситуація, коли в жесті присутнє перекривання між пальцями або фронтальне направлення пальців до співрозмовника тощо. В силу цього при розпізнаванні жесту виникають труднощі для ідентифікації фронтально направлених пальців або областей перекривання пальців.

У роботі [6] пропонується використовувати стереозір для вирішення даної проблеми. Стереозір включає в себе два процеси: співставлення деталей, які спостерігаються двома чи більше камерами, і відтворення їх тривимірного прообразу [7]. Так, співставлення деталей що фіксуються очима людини, та використання їх відмінностей надає можливість покращити уявлення про глибину об’єкта. У свою чергу, відтворення тривимірного прообразу точок можна знайти як точку перетину променів, що проходять через зазначені точки та центри відповідних діафрагм камер.

Метою даної роботи є використання контурного аналізу для підвищення ефективності розпізнавання жестів, які подібні за відтворенням жестів, у яких присутнє фронтальне розміщення пальців до співрозмовника, а також визначення межі пальців, прижатих один до одного, та ідентифікація області перекриття пальців.

У роботі запропоновано ряд алгоритмів для ідентифікації окремих елементів жестів, у яких використовується попіксельний аналіз зображення, і елементів дактиля. Даний підхід надає можливість правильно розпізнавати дактиль і уникнути певних труднощів при розпізнаванні дактиля на різних фокусних відстанях. Зокрема, наведені алгоритми отримання контурної інформації окремих елементів дактильної мови, які сприятимуть підвищенню рівня правильного розпізнавання складових дактильної мови.

Класифікація дактилем

Системи розпізнавання об’єктів різної природи, як відомо, повинні володіти множиною ознак, за якими можна класифікувати об’єкти щодо різноманітних груп. Зауважимо, що розпізнавання жестів можна визначити як відношення вхідних даних до певного класу образів за допомогою виділення істотних ознак або властивостей, які відділяють вхідні дані від загальної маси неістотних деталей. У свою чергу, під класом образів розуміється певна категорія, яка визначається рядом ознак, спільних для всіх її елементів [8]. Тому запропонована у роботі [9] схема класифікації жестів дактильної мови потребує певної деталізації та розширення (рис. 1).

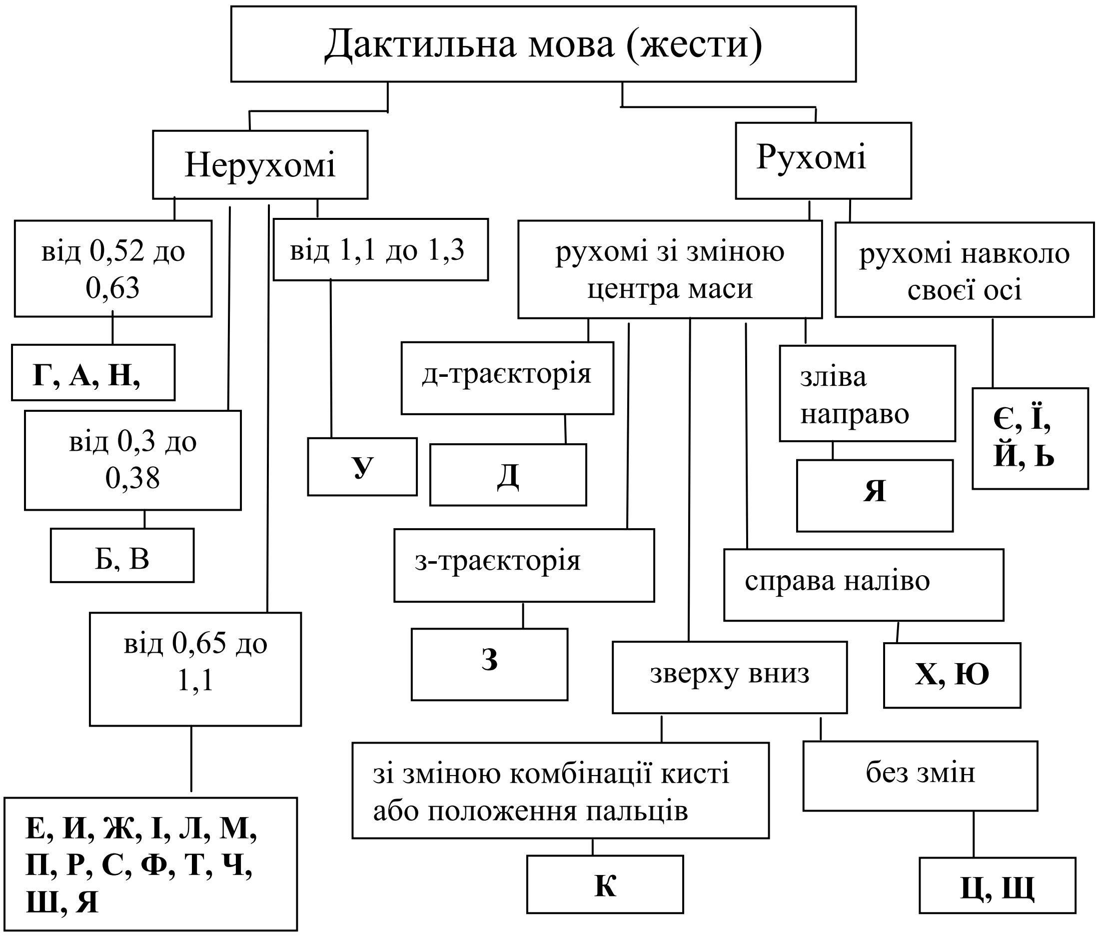


Рисунок 1 – Класифікація жестів дактильної мови

Так, жести, які відображають букви алфавіту, в загальному випадку можна поділити на дві групи – рухомі і нерухомі. До рухомих жестів відносяться жести, які відображають такі букви, як: Д, Є, З, Ї, Й, К, Х, Ц, Щ, Ю, Я, Ь. До нерухомих – А, Б, В, Г, Е, И, Ж, І, Л, М, Н, О, П, Р, С, Ф, Т, У, Ч, Ш, Я (букву Я можна відтворювати з рухом або без нього, тому така буква може належати до двох груп). Загалом, жест на відеозображенні розглядається як послідовність кадрів, на яких відбувається зміна форми кисті руки.

Слід зазначити, що група з рухомих жестів поділяється на групу з жестів, що відтворюються зі зміною центра координат, зокрема Д, З, К, Х, Ц, Щ, Ю, Я, і на групу жестів, що відтворюються навколо своєї осі, а саме – Є, Ї, Й, Ь. Жести, що

відображаються зі зміною центра координат, поділяються на жести, що відтворюються з різними траєкторіями руху: д-траєкторія – Д, з-траєкторія – З, зверху вниз – К, Ц, Щ, справа наліво – Х, Ю, зліва направо – Я. У свою чергу, група жестів, що віображається зверху вниз, поділяється на жести, що відтворюються зі зміною комбінації кисті, або положення пальців, наприклад, К, та без змін комбінації кисті, наприклад, Ц, Щ.

Група нерухомих жестів розподіляється за відношенням ширини до висоти прямокутної області навколо кисті, що показує жест. Відношення ширини до висоти складає: від 0,52 до 0,63 – Г, А, Н, О; від 0,3 до 0,38 – Б, В; від 0,65 до 1,1 – Е, И, Ж, І, Л, М, П, Р, С, Ф, Т, Ч, Ш, Я; від 1,1 до 1,3 – У (рис. 1).

Розпізнавання жестів

З метою розпізнавання елементів дактиля були запропоновані окремі алгоритми для ідентифікації елементів жестів, у яких використовується попіксельний аналіз зображення та елементів жесту [9]. На вхід даним алгоритмам подається область точок однакового кольору, що відносяться до жесту (рис. 2). Разом з тим існують жести, що є подібними за відтворенням, у зв'язку з чим їх важко визначити, використовуючи запропоновані алгоритми.

Для вирішення даної проблеми пропонується використовувати дві веб-камери, об'єктиви яких розташовані перпендикулярно до області, де відтворюється жест (рис. 3). Але такий підхід викликає додаткові просторові обмеження для користувача, тому для отримання більш інформативних вхідних даних пропонується будувати контури кисті і пальців, що відтворюють жест.



Рисунок 2 – Зразок вхідних даних для алгоритмів, котрі описані в роботі [9]

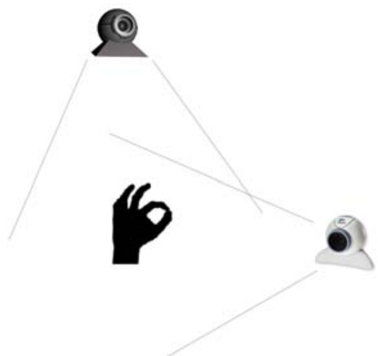


Рисунок 3 – Розташування камер для розпізнавання жесту

Так, нехай $f(x,y)$ – зображення, на якому відтворено жест. Для визначення наявності контуру в деякій області зображення можна використати градієнт яскравості:

$$grad(f(x,y)) = \nabla f(x,y),$$

де градієнт $grad(f(x,y))$ – це вектор, який напрямлений у напрямку найбільш швидкого зростання функції $f(x,y)$.

Зобразимо градієнт у наступному вигляді [7]:

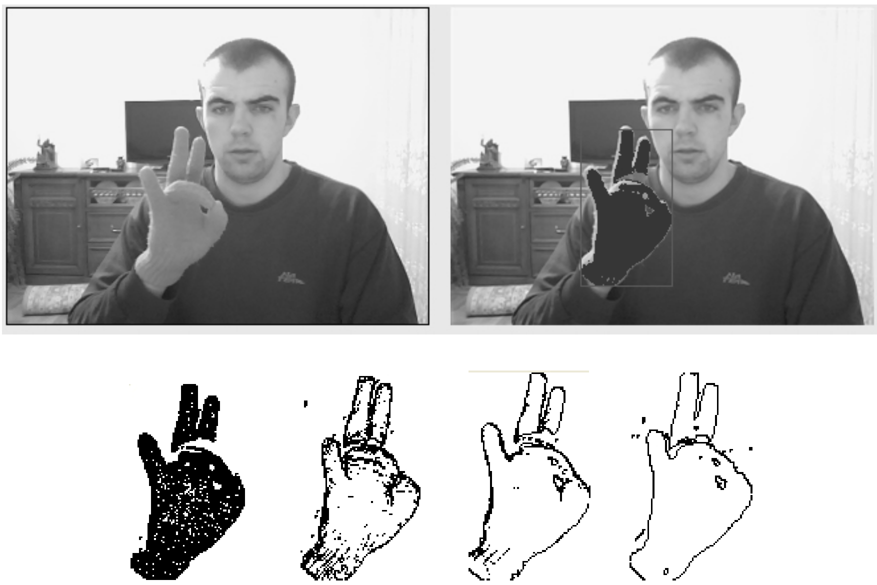
$$(\nabla^2 f)(x,y) = \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2, \ |\nabla f(x,y)| = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}.$$

У випадку цифрової обробки інформації частинні похідні $\frac{\partial f(x,y)}{\partial x}$ і $\frac{\partial f(x,y)}{\partial y}$ можна обрахувати наступним чином:

$$\frac{\partial f(x,y)}{\partial x} = f(x,y) - f(x-1,y), \ \frac{\partial f(x,y)}{\partial x} = f(x,y) - f(x,y-1).$$

Якщо градієнт більше заданого порогу k , то дана точка відноситься до контуру. Поріг задається користувачем. При різних порогах отримуються різні контури (рис. 4).

На основі отриманих зображень будується бібліотека еталонів для порівняння. Під час виконання жесту система визначає положення руки, обробляє зображення запропонованим алгоритмом. Отримане зображення порівнює з зображеннями з еталона і визначає, до якого класу відноситься буква.



$k = 0$ $k = 3$ $k = 5$ $k = 50$

Рисунок 4 – Відтворення букви «Н». Побудова контуру жесту з різними порогами

Розглянемо ще один спосіб для визначення контурів зображення, а саме – оператор Собеля. Результатом застосування оператора Собеля в кожній точці зображення є або вектор градієнта яскравості в даній точці, або його норма. Оператор Собеля ґрунтується на згортці зображення невеликими цілочисельними фільтрами у вертикальному і горизонтальному напрямі, а тому його відносно легко обрахувати.

Оператор Собеля використовує ядра 3*3, що дозволяє звертати вхідне зображення для обчислення наближених значень похідних по горизонталі і по вертикалі. Нехай A – вхідне зображення, а G_x і G_y – два зображення, де кожна точка містить приблизні значення похідних по x і по y. Значення G_x і G_y обраховуються наступним чином: $G_x = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * A$, $G_y = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} * A$, де * означає двомірну операцію згортки.

Координати x зростають «направо», а y зростають «вниз». Використовуючи отримані наближені значення похідних, в кожній точці зображення наближене значення величини градієнта можна обрахувати наступним чином: $G = \sqrt{G_{x^2} + G_{y^2}}$.

Якщо градієнт більше заданого порога k , то дана точка відноситься до контуру. Поріг задається користувачем.

Роботу запропонованого алгоритму представлено на рис. 5.

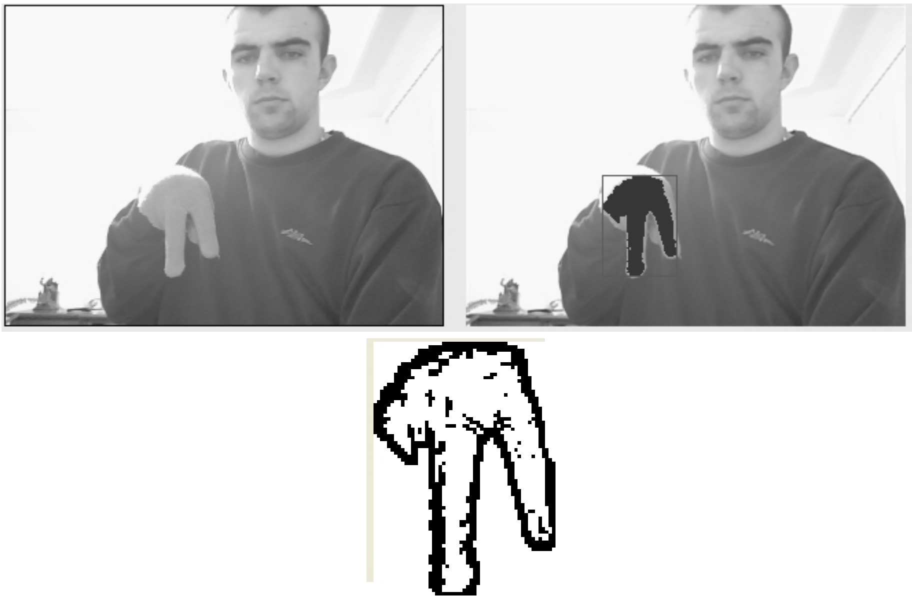
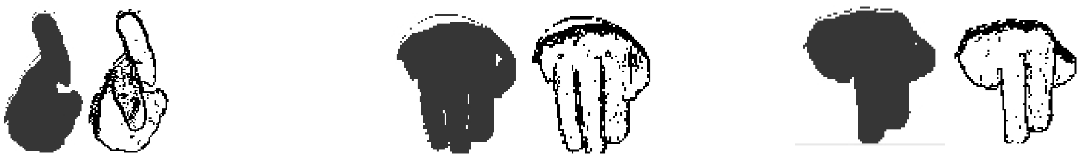


Рисунок 5 – Використання оператора Собеля

Таким чином, запропоновані методи надають можливість чіткіше розпізнавати букви, в комбінації яких присутнє фронтальне направлення пальців до співрозмовника, наприклад, Г (рис. 6), а також букви, котрі подібні за відтворенням, зокрема букви Н, О, Р або ж букви Т, П, що є також подібними між собою (рис. 6).



Г **Т** **П**

Рисунок 6 – Вхідні дані для різних методів

Результати експериментальних досліджень

Для тестування розробленої системи були проведені експерименти, в яких приймали участь люди різного віку з різними розмірами руки, а тому вищенаведені алгоритми показали досить хороші результати при розпізнаванні жестів, що відтворюються такими користувачами на різних фокусних відстанях.

У ході експерименту людина відтворює жест по п’ятдесят разів на різних фокусних відстанях і підраховується кількість разів, коли система правильно визначила жест. У табл. 1 і 2 наведено результати експериментів, що проводились без використання та з використанням контурного аналізу.

Таблиця 1 – Результати роботи системи при розпізнаванні букви Т без використання контурного аналізу

	Фокусні відстані від камери до руки					
	від 30 см до 40 см		від 40 см до 60 см		від 60 см до 90 см	
Особи, які відтворювали жест	Кількість правильно розпізнаних жестів з 50 разів відтворення	Відсоток правильно розпізнаних жестів з 50 разів відтворення (%)	Кількість правильно розпізнаних жестів з 50 разів відтворення	Відсоток правильно розпізнаних жестів з 50 разів відтворення (%)	Кількість правильно розпізнаних жестів з 50 разів відтворення	Відсоток правильно розпізнаних жестів з 50 разів відтворення(%)
	Буква Т					
Підліток 14 років	30	60	29	58	30	60
Підліток 16 років	33	66	33	66	32	64
Чоловік 23 років	30	60	31	62	34	68
Жінка 25 років	35	70	34	68	33	66
Чоловік 32 років	31	62	30	60	33	66
Чоловік 46 років	32	64	31	62	32	64
Чоловік 50 років	33	66	32	64	36	72
Середнє значення		**64**		**62,9**		**65,7**

Таблиця 2 – Результати роботи системи при розпізнаванні букви Т з використанням контурного аналізу

	Фокусні відстані від камери до руки					
	від 30 см до 40 см		від 40 см до 60 см		від 60 см до 90 см	
Особи, які відтворювали жест	Кількість правильно розпізнаних жестів з 50 разів відтворення	Відсоток правильно розпізнаних жестів з 50 разів відтворення (%)	Кількість правильно розпізнаних жестів з 50 разів відтворення	Відсоток правильно розпізнаних жестів з 50 разів відтворення (%)	Кількість правильно розпізнаних жестів з 50 разів відтворення	Відсоток правильно розпізнаних жестів з 50 разів відтворення (%)
	Буква Т					
Підліток 14 років	40	80	36	72	36	72
Підліток 16 років	38	76	37	74	36	72
Чоловік 23 років	39	78	36	72	35	70
Жінка 25 років	37	74	36	72	38	76
Чоловік 32 років	37	74	38	76	40	80
Чоловік 46 років	36	72	40	80	39	78
Чоловік 50 років	39	78	35	70	36	72
Середнє значення		**76**		**73,7**		**74,3**

Отже, з наведених результатів експериментів можна зробити висновок, що використання запропонованого методу збільшує кількість вірного розпізнавання жесту, який відтворюється особами з різними розмірами руки на різних фокусних відстанях. Так, результативність розпізнавання зросла в середньому від 64% до 75% відтворюваних жестів. Наведені результати вказують на ефективність використання запропонованого методу.

Висновки

У статті було розширено загальну класифікацію жестів дактильної мови. Описано алгоритми для більш інформативного збору ознак жесту, що розпізнається, зокрема алгоритми виділення контуру об'єкта. Отримані результати, після застосування запропонованих алгоритмів, надають можливість чіткіше розпізнавати жести, техніка виконання яких є подібною. Також використання контурного аналізу дає можливість більш якісно розпізнати елементи дактиля, у відтворення яких входить фронтальна орієнтація пальців до співрозмовника, чітко визначити межу між пальцями, які прижаті один до одного тощо.

В подальшому, на основі проведених експериментів і отриманих результатів, а також при використанні методів скелетизації, розроблятимуться алгоритми для ідентифікації жестів за допомогою скелетної моделі руки. Ідентифікація жесту за допомогою скелетної моделі дасть можливість більш якісно розпізнати жести, а також розробити синтезатор елементів дактильної мови.

Література

1. Fang G. Large-vocabulary continuous sign language recognition based on transition-movement models / G. Fang, W. Gao, D. Zhao // IEEE Trans. on Systems, Man, and Cybernetics . – Jan. 2007. – Vol. 37(1). – P. 1-9.
2. Real time large vocabulary continuous sign language recognition based on op/viterbi algorithm / G. Yao, H. Yao, X. Liu, F. Jiang // ICPR. – Aug. 2006. – Vol. 3. – P. 312-315.
3. Vogler C. A framework for recognizing the simultaneous aspects of American sign language / C. Vogler, D. Metaxas // CVIU. – Mar. 2001. – Vol. 81(3). – P. 358-384.
4. Hidden conditional random fields for gesture recognition / [S.B. Wang, A. Quattoni, L.-P. Morency, D. Demirdjian, T. Darrell] // CVPR. – June 2006. – Vol. 2. – P. 1521-1527.
5. Ong S. Automatic sign language analysis: A survey and the future beyond lexical meaning / S. Ong, S. Ranganath // PAMI. – June 2005. – № 27(6). – P. 873-891.
6. Smoothed Disparity Maps for Continuous American Sign Language Recognition / Philippe Dreuw, Pascal Steingrube, Thomas Deselaers, and Hermann Ney // Pattern Recognition and Image Analysis. – 2009. – P. 24-31.
7. Форсайт Д. Компьютерное зрение. Современный подход / Д. Форсайт, Ж. Понс. – М. : Вильямс, 2004. – 926 с.
8. Ту Дж. Принципы распознавания образов / Дж. Ту, Р. Гонсалес. – М. : Мир, 1978. – 411с.
9. Крак Ю.В. Технологія розпізнавання елементів дактильно-жестової мови / Ю.В. Крак, Д.В. Шкільнюк // Штучний інтелект. – 2009. – № 3. – С. 564-572.

Ю.В. Крак, Д.В. Шкильнюк

Анализ элементов дактильного языка

В статье рассматриваются отдельные методы повышения качества распознавания элементов дактильного языка, в частности конкретизирована схема классификации жестов. Использование и реализация данных методов способствует качественному распознаванию составляющих дактильного языка.

Yu.V. Krak, D.V. Shkilnyuk

Analysis of Elements of Finger-Sing Language

This article discusses some methods to improve the quality of recognition elements of deaf language, in particular specified the classification scheme deaf-sign language. The introduction and implementation of these methods contributes to quality recognition composite deaf-language.

Стаття надійшла до редакції 19.07.2010.